Available online at www.sciencedirect.com
SCIENCE DIRECT®
J. Biochem. Biophys. Methods 55 (2003) 233–240
JOURNAL OF biochemical and biophysical methods
www.elsevier.com/locate/jbbm

Expression of soluble recombinant proteins in a cell-free system using a 96-well format

Didier Busso[a,*], Rosalind Kim[b], Sung-Hou Kim[a,b]

[a] *Department of Chemistry, University of California, Berkeley, 314 Calvin Lab MS5230, CA 94720-5230, USA*
[b] *Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

Received 23 January 2003; accepted 24 February 2003

Abstract

For structural and functional genomics programs, new high-throughput methods to obtain well-expressing and highly soluble proteins are essential. Here, we describe a rapid procedure to express recombinant proteins in an *Escherichia coli* cell-free system using a 96-well format. The identification of soluble proteins is performed by the Dot Blot procedure using an anti-His tag antibody. The applications and the automation of this method are described.

Keywords: Cell-free; Screening; Expression; Solubility; Automation; Recombinant proteins

1. Introduction

During the last decade, many genomes from both prokaryotes and eukaryotes have been completely sequenced and more are under way (http://www.tigr.org/tdb/). Based on sequence analysis, a large fraction of genes has unknown cellular and/or molecular function. One major challenge is to assign biological function and to elucidate the mechanism of action of those open reading frames. The three-dimensional structure of a protein can often provide functional clues, primarily by detecting structural homology with a protein of known function even when sequence homology is low [1]. To face this challenge, several centers worldwide are undertaking structural genomics initiatives ([2]; see also (http://www.rcsb.org/pdb/strucgen.html#Worldwide).

* Corresponding author. Tel.: +1-510-486-4332; fax: +1-510-486-6059.
E-mail address: djbusso@lbl.gov (D. Busso).

0165-022X/03/$ - see front matter
doi:10.1016/S0165-022X(03)00049-6

For structural studies, high yields of soluble recombinant proteins are required. Due to fast growth, easy handling and low cost, *Escherichia coli* is the principal expression system of choice [3,4]. Nevertheless, recombinant proteins produced in *E. coli* often accumulate as insoluble aggregates. Changing parameters such as temperature, additives, induction conditions or adding fusion partners may alter the behavior of the recombinant proteins. Therefore, the development of effective and fast-screening methods for expression and solubility is necessary. An alternative to in vivo expression screening is the less time-consuming cell-free expression system.

Here, we present the results obtained from 24 open reading frames (ORF) of unknown function from different microorganisms. In order to screen different variables that may interfere with solubility, we expressed the recombinant proteins with either an N-terminal or a C-terminal 6 histidine tag using a cell-free expression system at two temperatures (25 and 30 °C). The screening was done in a 96-well format and the expression-solubility characterization was determined by performing Dot Blots using an antibody directed against the histidine tag.

We designed a rapid method that allows (i) the characterization of soluble candidates from a large number of genes or from a large number of variants, (ii) complete automation of the process and (iii) an immediate scale-up of protein expression for the selected candidates.

2. Materials and methods

2.1. Expression vector construction

The genes of interest (Table 1) were amplified by polymerase chain reaction (PCR) from the corresponding genomic DNA using a forward primer introducing an *Nde*I site and a reverse primer introducing a Stop codon and a *Bam*HI site. The PCR products were inserted using the ligation activity of topoisomerase into a TOPO vector (Invitrogen, Carlsbad, CA). At this stage, PCR products were confirmed by DNA sequencing. DNA inserts encoding the proteins of interest were prepared by digestion using *Nde*I and *Bam*HI (New England Biolabs, Beverly, MA) followed by a gel purification step using the StrataPrep® DNA gel extraction kit (Stratagene, La Jolla, CA). These *Nde*I–*Bam*HI inserts were originally prepared for other cloning procedures.

To express the recombinant proteins under control of the T7 promoter with an N-terminal 6 histidine tag (Nt His tag), the *Nde*I–*Bam*HI inserts were subcloned in frame into the pIVEX2.4b-*Nde* vector (Roche Applied Science, Mannheim, Germany) digested with the corresponding enzymes. In order to express the recombinant proteins under control of the T7 promoter with a C-terminal 6 histidine tag (Ct His tag), the removal of the stop codon was necessary. Thus, the genes of interest were amplified by PCR from the corresponding recombinant pIVEX2.4b-*Nde* vector using the universal T7 promoter primer (5′-TAATACGACTCACTATAGGG-3′) and a gene specific reverse primer introducing an *Xho*I site and removing the stop codon. After quantification, 1 μg of PCR product was digested by *Nde*I and *Xho*I (New England Biolabs) and gel purified. The restricted PCR products were subsequently subcloned in frame into the pIVEX2.3-MCS vector (Roche Applied Science) digested with the corresponding enzymes.

Table 1
List of proteins selected for in vitro expression-solubility screening

	Protein	Organism	GI Number	M_r (kDa)
1	AA967	*Aquifex aeolicus*	2983573	36.2
2	BS967	*Bacillus subtilis*	2633961	35.7
3	MP967	*Mycoplasma pneumoniae*	1673967	36.6
4	BS217	*Bacillus subtilis*	2633884	16.5
5	MG217	*Mycoplasma genitalium*	3844824	17.8
6	MP217	*Mycoplasma pneumoniae*	1674217	16.3
7	BS676	*Bacillus subtilis*	2635300	19.1
8	BS742	*Bacillus subtilis*	2632368	20.1
9	BS994	*Bacillus subtilis*	2633216	18.6
10	EC731	*Escherichia coli*	1790387	11.7
11	EC836	*Escherichia coli*	1787757	15.1
12	MG222	*Mycoplasma genitalium*	3844817	44.6
13	TM142	*Thermatoga maritima*	4981173	31.2
14	TM915	*Thermatoga maritima*	4981625	15.8
15	MP003	*Mycoplasma pneumoniae*	1674003	13.6
16	MP004	*Mycoplasma pneumoniae*	1674004	16.8
17	MP034	*Mycoplasma pneumoniae*	1674034	16.2
18	MP182	*Mycoplasma pneumoniae*	1674182	13.4
19	MP235	*Mycoplasma pneumoniae*	1674235	17.7
20	MP278	*Mycoplasma pneumoniae*	1674278	29.0
21	MP865	*Mycoplasma pneumoniae*	1673865	31.0
22	MP883	*Mycoplasma pneumoniae*	1673883	15.4
23	MP920	*Mycoplasma pneumoniae*	1673920	40.5
24	MP958	*Mycoplasma pneumoniae*	1673958	22.4

The protein name is based on the two letters corresponding to the DNA of the organism the PCR was done from and the three digits correspond to the last three digits of the *Mycoplasma pneumoniae* GI number. This means that the two proteins with the same three digits and different letters represent two homologues from different organisms of the corresponding *M. pneumoniae* open-reading frame. The first column gives the protein order used in Fig. 1. M_r: molecular mass.

All DNA preparations of the recombinant expression vectors (pIVEX2.4b-*Nde* and pIVEX2.3-MCS) were made using the QIAprep spin miniprep kit (Qiagen, Valencia, CA). DNA concentrations, either from mini preparations or from PCR reactions, were measured using a Shimadzu spectrophotometer.

2.2. In vitro protein expression in 96-well plates

In vitro screening for protein expression and solubility was carried out using the cell-free protein synthesis system developed in the Genomic Science Center (Institute of Physical and Chemical Research (RIKEN), Yokohama, Japan) by the team of Dr. Shigeyuki Yokoyama. Basically, the cell-free protein synthesis system consists of a coupled transcription–translation reaction from a recombinant DNA source using an *E. coli* S30 cell extract, T7 RNA polymerase and low molecular weight substrates [5,6]. The S30 cell extract was prepared by modification of the procedure of Pratt [7] from the BL21(DE3) codon plus RIL strain (Stratagene) as described elsewhere [8,9]. Briefly, the *E. coli* cells were resuspended in buffer A (60 mM potassium acetate, 10 mM

tris(hydroxymethyl)aminomethane (*Tris*)-acetate pH 8.2, 14 mM magnesium acetate, and 1 mM dithiothreitol (DTT) plus 7.2 mM β-mercaptoethanol and centrifuged (16,000 × *g*, 30 min, 4 °C). The cells were disrupted by microfluidization (Microfluidics, Newton, MA) in buffer A and centrifuged (30,000 × *g*, 30 min, 4 °C). Pre-incubation buffer (300 mM *Tris*-acetate pH 8.2, 10 mM magnesium acetate, 10 mM ATP, 80 mM phosphoenol pyruvate (PEP), 5 mM DTT, 40 μM amino acid mix) was added to the cell suspension in the presence of pyruvate kinase and then dialyzed against buffer A. The T7 RNA polymerase was purified, as previously described [10,11].

For expression in microtiter plates, the reaction mix (50 μl) consisted of 58 mM Hepes–KOH, pH 7.5, 4% polyethylene glycol (PEG) 8000, 200 mM potassium glutamate, 1.8 mM DTT, 1.2 mM ATP, 0.8 mM each of CTP, GTP and UTP, 0.64 mM 3′, 5′-cyclic AMP, 35 μg/ml folinic acid, 27.5 mM ammonium acetate, 80 mM creatine phosphate, 0.25 mg/ml creatine kinase, 175 μg/ml *E. coli* total tRNA, 0.05% sodium azide, 9.6 mM magnesium acetate, 1.5 mM each of the 20 amino acids, 0.3 μl T7 RNA polymerase (800 U/μl) and 15 μl S30 extract. The reaction mix was dispensed into a MicroAmp® optical 96-well reaction plate (PE Applied Biosystems, Foster City, CA). After addition of 150 ng of DNA template (pIVEX2.4b-*Nde* and pIVEX2.3-MCS recombinant expression vectors), the reactions were incubated for 4 h either at 25 or 30 °C in a Gene-amp PCR system 9700 thermocycler (PE Applied Biosystems). After incubation, the soluble proteins were obtained by centrifuging the total reaction mix in the microplate at 3,000 × *g* for 20 min at 4 °C.

2.3. Dot Blot procedure

To test for recombinant protein expression, 1 μl of total protein extract and soluble proteins were spotted onto Hybond™ ECL™ nitrocellulose membrane (Amersham Pharmacia Biotech, Piscataway, NJ). The prepared membranes were air dried and then incubated with anti-His monoclonal mouse antibody (Amersham Pharmacia Biotech) diluted 1:5000 in 3% powdered milk for 1 h. Membranes were then washed three times for 10 min in *Tris*-buffered saline Tween (TBST: 50 mM *Tris*–HCl, pH 7, 250 mM NaCl, 0.05% Tween 20). The secondary antibody, anti-mouse IgG-horseradish peroxidase conjugate (Amersham Pharmacia Biotech) was applied at 1:2000 dilution in TBST + 3% BSA for 20 min. After washing three times for 10 min in TBST, blots were developed with 3,3′,5,5′-tetramethylbenzidine (TMB) solution (Promega, Madison, WI).

3. Results and applications

3.1. Screening for conditions expressing soluble protein

Twenty-four ORFs encoding hypothetical proteins ranging from 11.7 to 44.6 kDa from various microorganisms (Table 1) were used to test a method to identify conditions allowing soluble expression in vitro in a 96-well format. Recombinant proteins carrying either an N-terminal 6 histidine tag (Nt His tag) or a C-terminal 6 histidine tag (Ct His tag) were expressed in a 50-μl reaction using a cell-free expression system at two temperatures

(25 and 30 °C). Total protein and soluble protein expression were analyzed by Dot Blots (Fig. 1a and b, respectively). The results obtained show that of the 24 ORFs, three exhibited low expression levels (BS967, EC836 and MP883) although these proteins were expressed at a detectable level in vivo. Optimizing the in vitro expression conditions may increase the

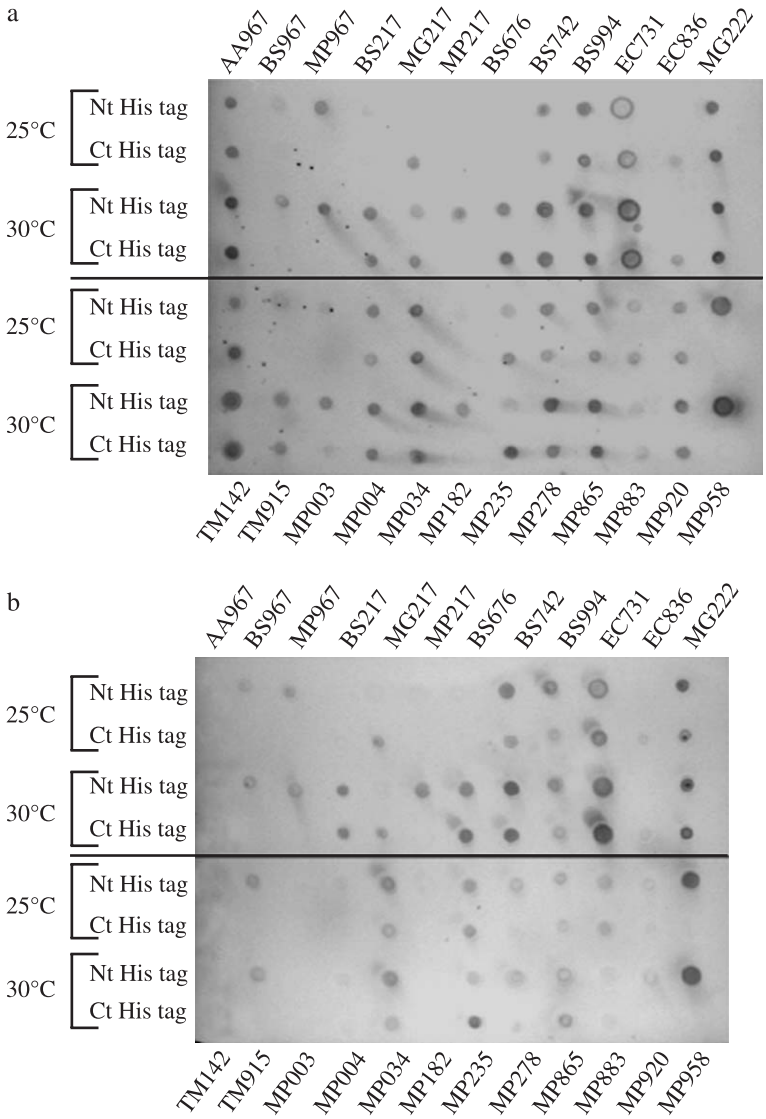

Fig. 1. Dot Blots of recombinant proteins expressed in vitro at 25 and 30 °C. Protein samples (1 μl) were spotted onto nitrocellulose membrane and probed for His-tagged proteins with an anti-His antibody (see Materials and methods). (a) Total proteins from the reaction mix. (b) Soluble protein from the centrifuged reaction mix. Nt His: N-terminal 6 histidine tag. Ct His: C-terminal 6 histidine tag.

expression level of these proteins. About 50% of the expressed proteins show detectable soluble expression in vitro correlating to data obtained in vivo (Busso, unpublished work). Moreover, the location of the histidine tag may affect not only solubility (see Nt His tag TM915-MP278 and Ct His tag MG217) but also the expression level (see Nt His tag BS967-MP967-MP217-MP003-MP182-MP958 and Ct His tag EC836) of recombinant proteins in vitro. Lowering the temperature did not seem to increase soluble expression of these recombinant proteins in vitro as compared to those in vivo [4]. In our experiment, only MP883 was expressed at a higher level at 25 °C than at 30 °C.

The present method described using the Dot Blot procedure for protein identification via a universal tag allows good qualitative detection of clones expressing soluble protein. The method has not been optimized for quantitative measurement of protein concentrations. However, as low as 10 ng of His tagged protein spotted may be detected using this procedure (see Material and methods). Nevertheless, since the in vitro expression reaction was performed in the presence of the same amount of plasmid DNA template, our procedure gives a good qualitative detection of clones expressing recombinant His-tagged proteins. The method allows a quick comparison of the level of expression by scanning the blots or by visual check.

Since it appears that most of the proteins behave similarly in vivo and in vitro ([8,12]; Yokoyama, RIKEN Genomic Sciences Center, personal communication and Busso, unpublished work), the cell-free expression system coupled with colorimetric detection of expressed proteins may be an alternative to the time-consuming in vivo system to identify quickly the constructs and parameters that yield soluble expression.

3.2. Automation

The experiments presented in this preliminary study were done using classical laboratory equipment. Nevertheless, the usage of multichannel pipettes and 96-well plates allow us to establish a procedure for automation using the Biomek 2000 robot (Beckman Coulter, Palo Alto, CA). Almost all cloning steps including PCR reactions, DNA restrictions, DNA cleanup, ligation and transformation as well as DNA mini-preparations are automated using commercial available kits. DNA quantification is performed in a microplate reader using a 96-well UV plate (Corning, Fountain Valley, CA). For the in vitro expression assay, appropriate dilution of plasmid DNA mini-preparations are added to the reaction mix already dispensed in a microtiter plate by the robot. The reaction plate is then moved to a Gene-amp PCR system 9700 thermocycler (PE Applied Biosystems) by the gripper tool. After incubation at the appropriate temperature, an aliquot of the total protein is transferred to a microtiter plate for further analysis. The soluble proteins are collected by filtering the total protein extracts through a 96-well filter plate (multiscreen-DV Durapore-0.65-m pore size-Millipore, Bedford, MA) to a second microtiter plate on a vacuum manifold [13]. Both total protein and the soluble fraction are spotted on nitrocellulose membrane by the robot.

It is also interesting to note that the cell-free protein expression system is also compatible with expression screening using a PCR product harboring all sequences required for in vitro expression as a template ([14]; Yokoyama, personal communication). This approach allows fast screening to select candidates compatible with soluble

expression. For further analysis or for large-scale expression, the blunt PCR products may be quickly inserted into a TOPO vector using the ligation activity of topoisomerase (Busso, unpublished results).

4. Conclusion

We have described a fast method to screen expression and solubility of recombinant proteins expressed in a cell-free system using a convenient 96-well format. The easy detection method using a universal tag allows a fast and reliable way to identify well-expressed soluble proteins. The automation of in vitro expression-solubility method will allow the screening of a large number of fusions in a large expression matrix (temperature, additives,. . .) and should be a very important tool for the ongoing efforts to characterize soluble proteins for structural studies. Finally, the scalability of the system allows large-scale protein production of the selected candidates yielding expression levels compatible with structural genomics expectations.

Acknowledgements

This work was supported by the National Institutes of Health GM 62412.

References

[1] Zarembiski TI, Hung L-W, Muller-Dieckmann H-J, Kim K-K, Yokota H, Kim R, et al. Structure-based assignment of the biochemical function of a hypothetical protein: a test case for structural genomics. Proc Natl Acad Sci U S A 1999;95:15189–93.
[2] Brenner SE. A tour of structural genomics. Nat Genet 2001;2:801–9.
[3] Baneyx F. Recombinant protein expression in *Escherichia coli*. Curr Opin Biotech 1999;10:411–21.
[4] Hannig G, Makrides SC. Strategies for optimizing heterologous protein expression in *Escherichia coli*. Trends Biotech 1998;16:54–60.
[5] Spirin AS, Baranov VI, Ryabova LA, Ovodov SY, Alakhov YB. A continuous cell-free translation system capable of producing polypeptides in high yield. Science 1988;242:1162–4.
[6] Kigawa T, Yokoyama S. A continuous cell-free protein synthesis system for coupled transcription-translation. J Biochem 1991;110:166–8.
[7] Pratt JM. In: Hames BD, Higgins SJ, editors. Transcription and Translation. Oxford, Washington DC: IRL Press 1984; p. 179-209.
[8] Kigawa T, Yamaguchi-Nunokawa E, Kodama K, Matsuda T, Yabuki T, Matsuda N, et al. Selenomethionine incorporation into a protein by cell-free synthesis. J Struct Funct Genomics 2002;2:29–35.
[9] Kigawa T, Muto Y, Yokoyama S. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. J Biol NMR 1995;6:129–34.
[10] Grodberg J, Dunn JD. *OmpT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol 1988;170:1245–53.
[11] Zawadzki V, Gross HJ. Rapid and simple purification of T7 RNA polymerase. Nucleic Acids Res 1991; 19:1948.
[12] Monchois V, Vincentelli R, Deregnaucourt C, Claverie J-M, Abergel C. Proficient target selection in structural genomics by in vitro protein expression. Biochemica 2002;1:22–3.

[13] Knaust RKC, Nordlund P. Screening for soluble expression of recombinant proteins in a 96-well format. Anal Biochem 2001;297:79–85.

[14] Rungpragayphan S, Kawarasaki Y, Imaeda T, Kohda K, Nakano H, Yamane T. High-throughput, cloning-independent protein library construction by combining single-molecule DNA amplification with in vitro expression. J Mol Biol 2002;318:395–405.